



$$E = (DC)^2$$

czyli jak można pracować w oparciu
o dwa lub więcej centra danych



Agenda:

- Rys historyczny
- Rozwój i problemy sieci
- Nowa struktura
- Podsumowanie



Start

- 1996 – nowa marka na rynku
Internet głównie na uczelniach, w firmach i domach – modemy
- jedna lokalizacja, jedno połączenie z Internetem



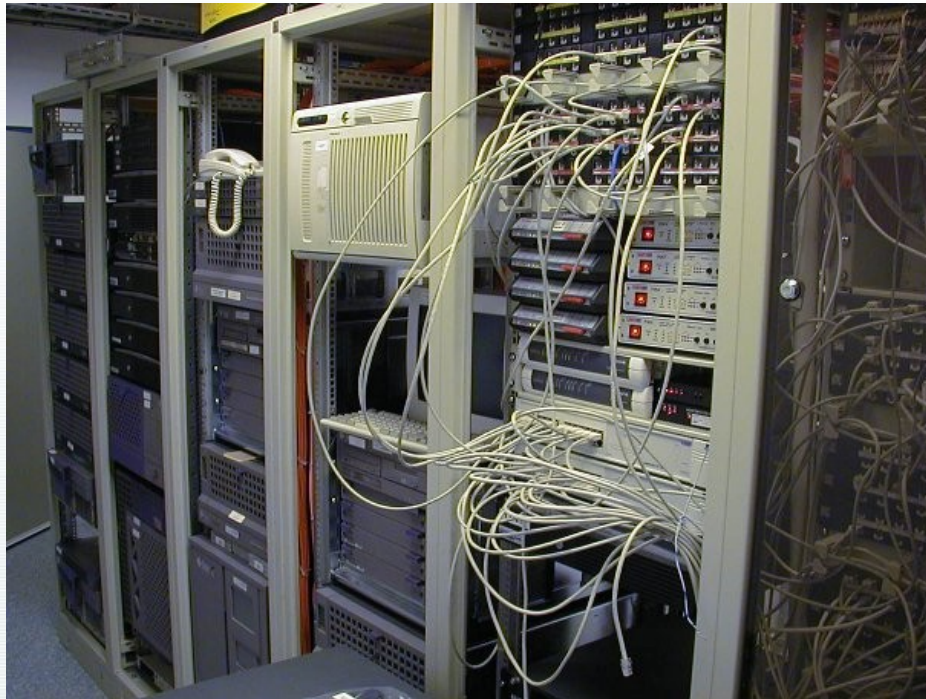
Rozwój 1996 - 1999

- kolejne zmiany lokalizacji, podyktowane wzrostem zapotrzebowania na miejsce dla ludzi i serwerów
- zasilanie - 12 kVA
- wzrost ruchu, kolejne łącza operatorskie



Pierwsza dedykowana serwerownia – AD 1999

- 10 szaf serwerowych
- 32 kVA mocy
- ponad 100 serwerów



- kolejne setki Mbps
- znów mało miejsca
- SPOF?



Dalszy rozwój...

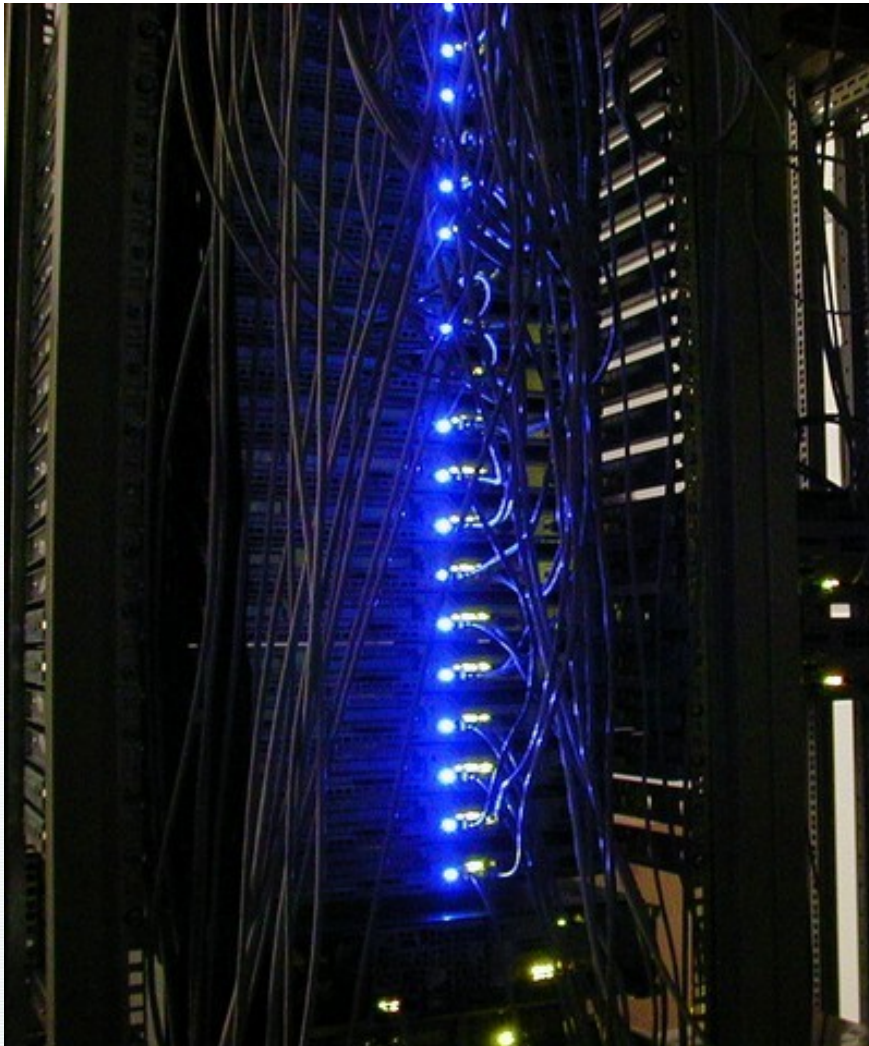


- 2002: wchodzimy do nowo wybudowanej serwerowni
- większe bezpieczeństwo
- więcej miejsca (dla ponad tysiąca serwerów)
- 320 kVA mocy

- rozwój usług, coraz większe przepustowości, usługi video
- więcej sprzętu, coraz większe zapewnienie serwerowni
- 2006 – pierwsze kolokacje



Onet.pl do AD 2008:



- dwie własne serwerownie
- więcej urządzeń w kolokacji niż „u siebie”
- budynek biurowy z serwerownią backoffice
- dodatkowe biura w Krakowie i Warszawie
- 1500 serwerów
- 8 dużych routerów, kilkadziesiąt switchy rackowych i drugie tyle w paczkach blade



Agenda:

- Rys historyczny
- **Rozwój i problemy sieci**
- Nowa struktura
- Podsumowanie



Rozwój sieci

AD 1999/2000:

- dwa VLANy
- niewielkie farmy serwerów
- jedna serwerownia

A później:

- coraz większe farmy serwerów
- setki współzależnych aplikacji
- kolejne lokalizacje
- kolejne łącza na świat



L2/L3?

- | w jednej serwerowni – nie ma problemu

- | w dwóch serwerowniach – rozciągnięcie L2 jest najbezpieczniejszym sposobem na wejście do drugiej serwerowni, potem L2 także daje radę
 - miejscami pojawia się L3
 - globalna zmiana na L3 jest trudna, ponieważ wymaga dużych zmian po stronie aplikacyjnej – łatwiej dokładać, niż modernizować

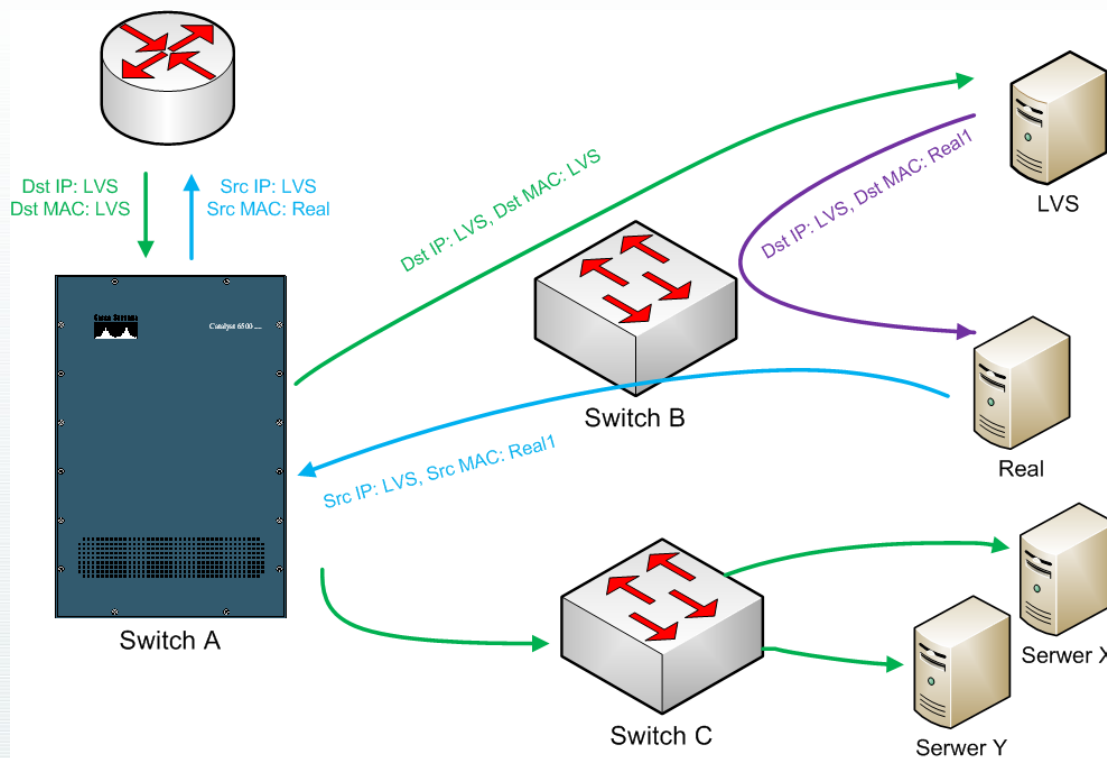
- | kolokacje i kolejne DC – tu zaczynają się schody

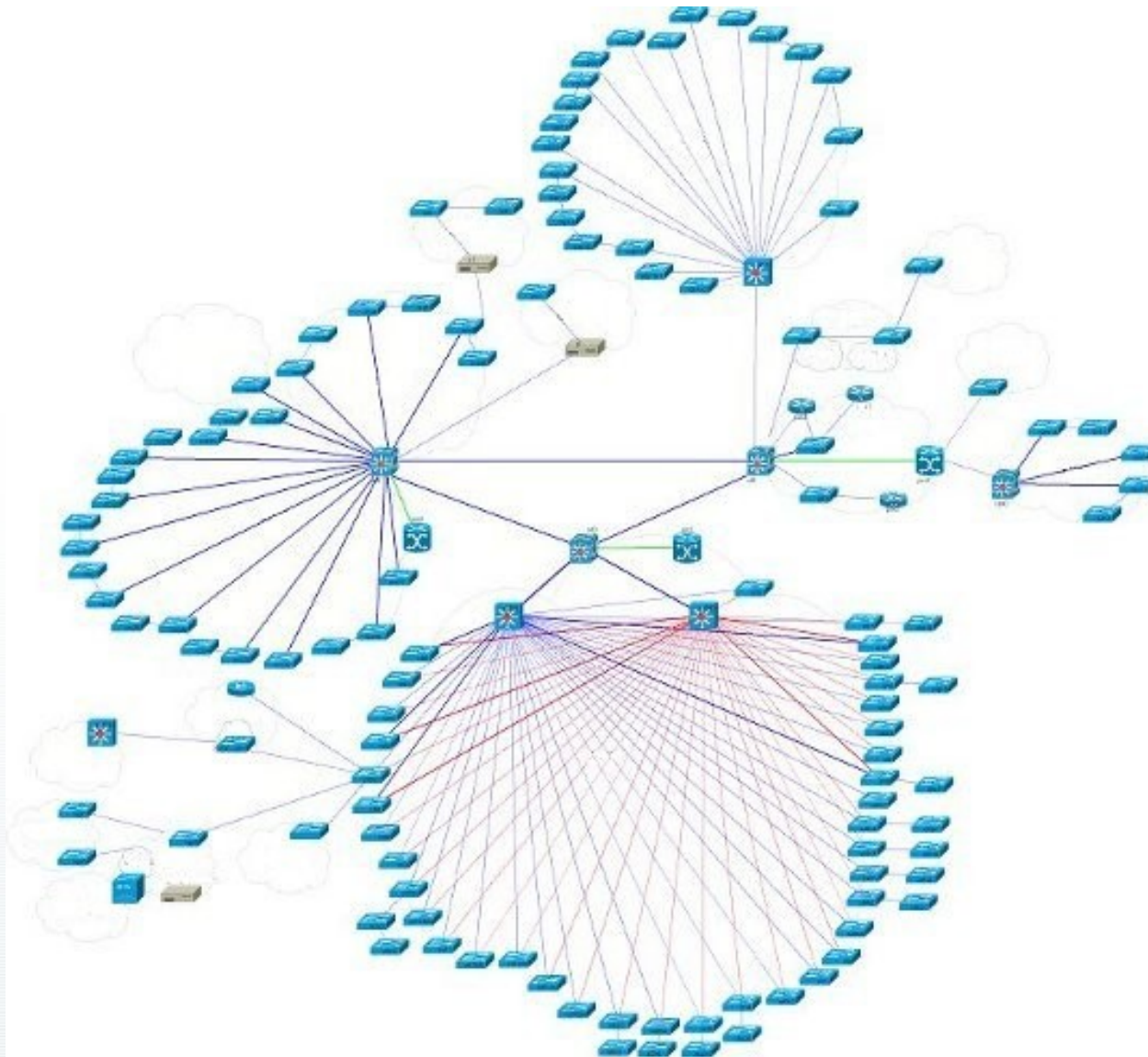


Problemy...

Rozrastanie się L2 i problemy skali:

- zmiany maski podsieci dla przyjęcia kolejnych serwerów
- sztormy broadcastowe
- MAC aging...





- L2: zbyt dużo segmentów LAN – nie skaluje się na kolejne DC



Problemy...

- tworzenie nowych usług w osobnych VLANach (L3) po pewnym czasie doprowadza do powstawania zbyt wielu VLANów (brak instancji Spanning Tree na niektórych switchach po przekroczeniu 128)

Oraz inne:

- Routery LAN – routerami WAN, awaria jednego urządzenia powoduje jednoczesne problemy z podawaniem się na zewnątrz oraz z ruchem wewnętrznym
- Rozwój Internetu - 256 tysięcy wpisów w tablicy routingu przestaje wystarczać



Co dalej?

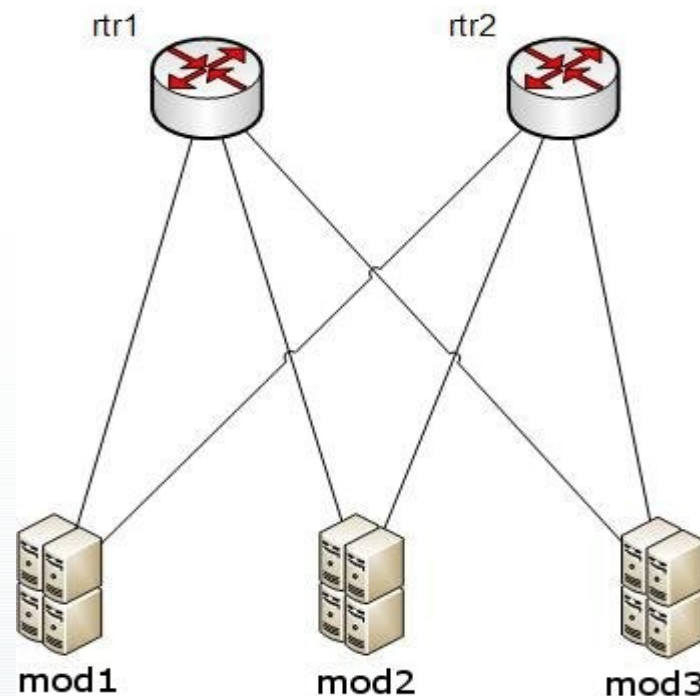


Zmiany...



Idea modułów - założenia

- autonomiczne systemy, mające od kilku do kilkudziesięciu urządzeń, obsługujące pewną pulę usług
- redundantne połączenia do warstwy agregacji
- komunikacja po L3, z wykorzystaniem routingu dynamicznego (czyli obsługiwane adresy IP mogą się zmieniać)
- samowystarczalne – możliwość pracy po utracie łączności z resztą sieci firmy (przy zachowaniu łączności ze światem), przy zachowaniu maksimum możliwej funkcjonalności
- możliwość przejęcia usług innego modułu



Zmiany do AD 2008...

1. Zmniejszanie liczby VLANów, między innymi przez stosowanie Private VLAN (jednocześnie odzyskujemy adresy L3)
3. Zwiększanie niezawodności L2/L3 – pełna redundancja ścieżek, przejście na Rapid Spanning Tree, HSRP w zdecydowanej większości VLANów
5. Wprowadzanie OSPF (równolegle do EIGRP)
7. Wyraźna migracja w kierunku L3 – wdrażanie zarysowanej wcześniej idei klocek (modułów)

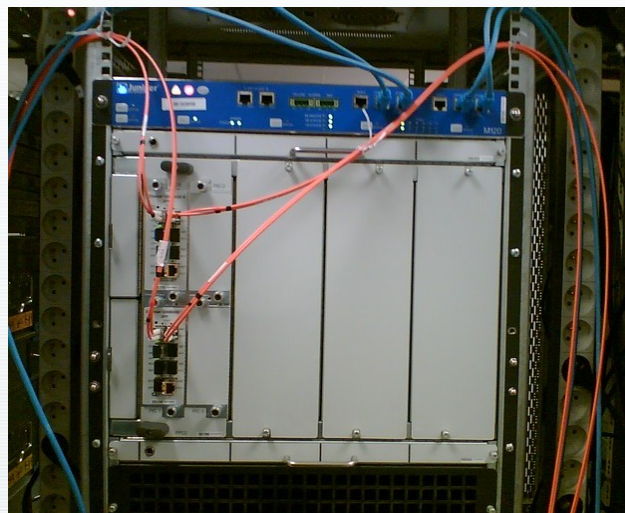


Zmiany cd...

1. Upgrade sprzętu:
nowe Supervisory do
dużych Catalystów,
w produkcji wszędzie
porty 1Gbps



6. Przeniesienie
peeringów BGP na
osobne routery -
wdrożenie Juniperów



Zmiany...



- | Ścisłejsze definiowanie funkcji poszczególnych serwerów i lepsze rozkładanie zadań (unikanie serwerów wielozadaniowych, tworzenia SPOFów)
- | Położenie większego nacisku na procedury DRP
- | Monitoring 24x7 – Administratorzy Pierwszej Linii Wsparcia (A1LW)
- | Testy w labie ↶

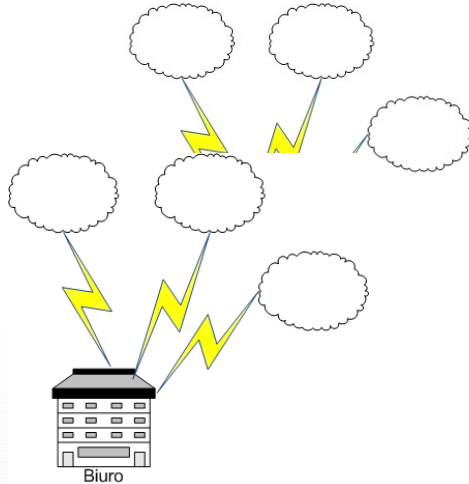


Agenda:

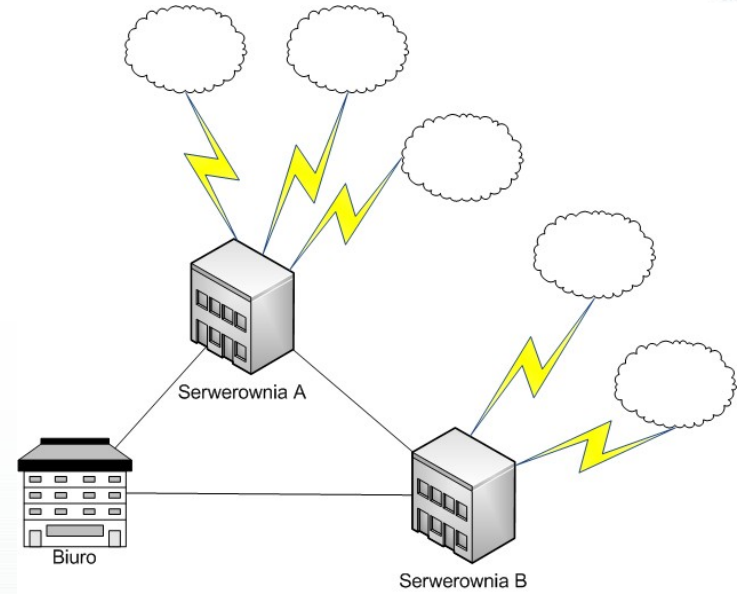

- Rys historyczny
- Rozwój i problemy sieci
- **Nowa struktura**
- Podsumowanie



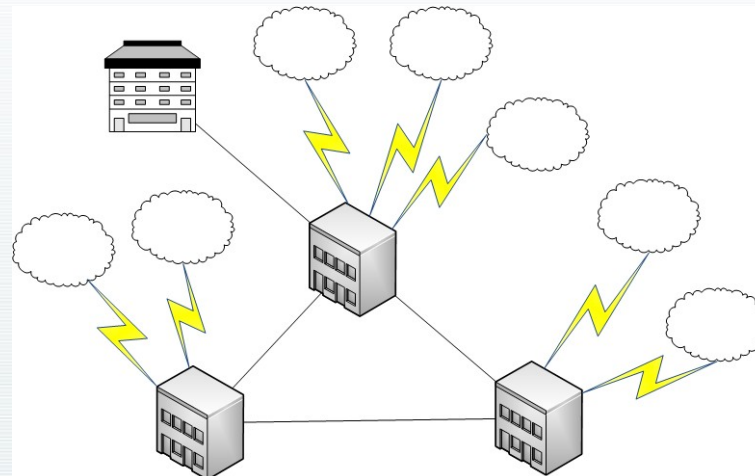
Ewolucja



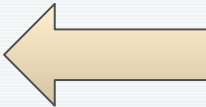
2002



2006



2009



?



2009: Onet Data Center



- trzy komory – trzy osobne serwerownie, połączone betonem i zasilaniem
- miejsce na pięć tysięcy serwerów
- ponad 2000 kVA mocy IT
- 5 MW mocy maksymalnej, normalny pobór to ponad 3 MW
 - to mniej więcej tyle, ile 7000 domów jednorodzinnych



Jak dobrze wykorzystać kilka DCs?

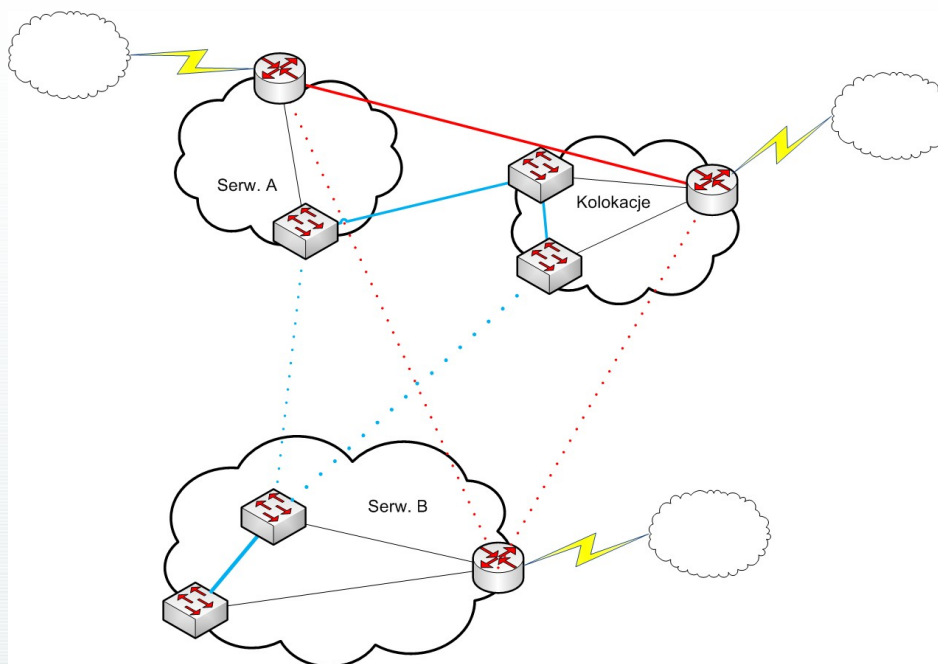
Czego chcieliśmy?

- pełna redundancja wewnątrz każdej z serwerowni
brak SPOFów spowodowanych wyjątkową rolą jednego z urządzeń
- łatwe przenoszenie usług/serwisów zarówno w ramach jednej serwerowni, jak i poza nią
sterowanie obciążeniem systemów podających content
- wykorzystanie w normalnej pracy wszystkich wyjść na świat
optymalne wykorzystanie łączy



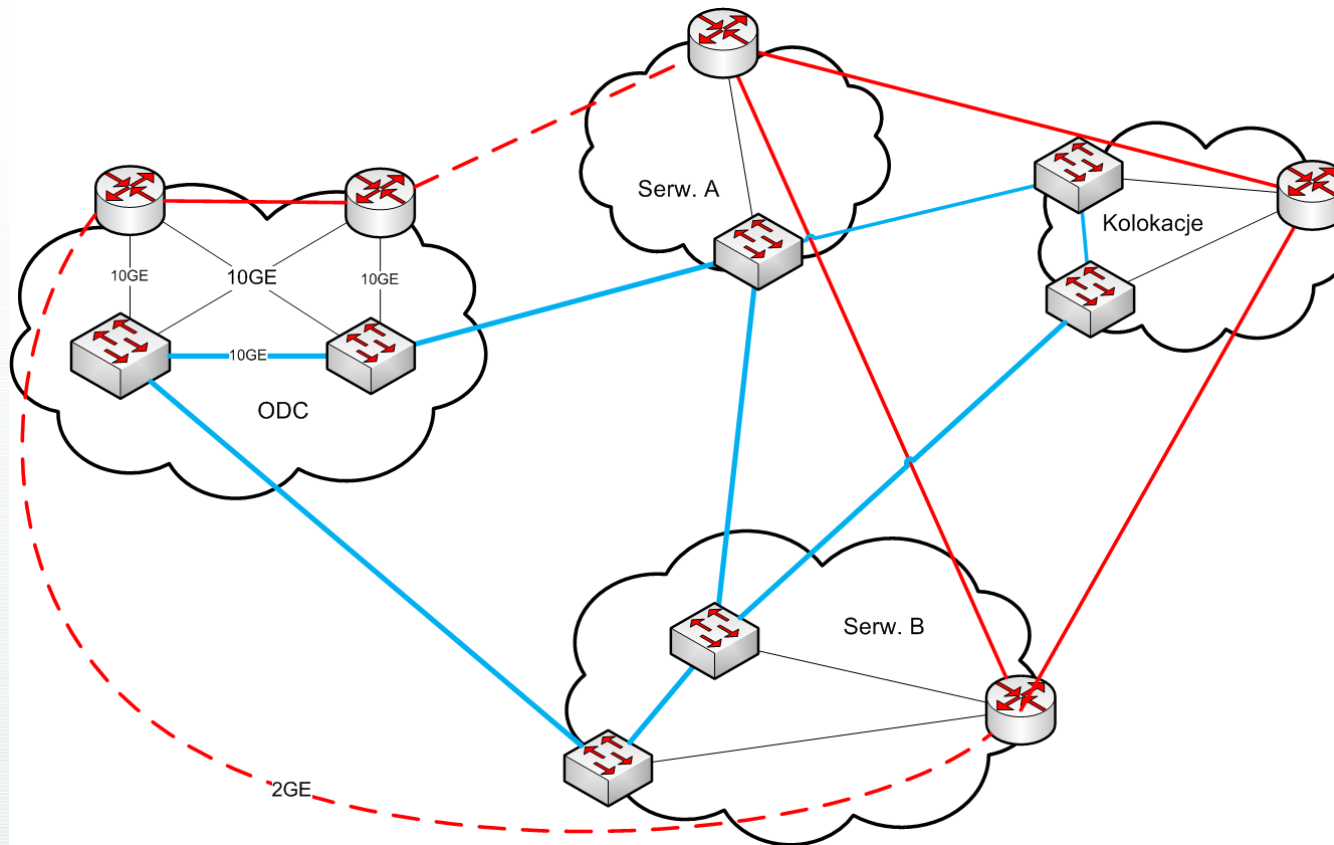
Jak dobrze wykorzystać kilka DCs?

- pełna redundancja serwerowni – skoro mamy peering BGP w każdej z nich, w przypadku dużej awarii każda z nich powinna móc serwować ruch samodzielnie



Jak dobrze wykorzystać kilka DCs?

- skalowalność: integracja z Onet Data Center i możliwość dołączania w przyszłości kolejnych DC



Projekt i wdrożenie - 2008/2009



- 2008 Q4 – 2009 Q1: 30 nocek w ciągu pięciu miesięcy
- mam nadzieję, że nie zauważyliście... ;)



Zmiany 2008/2009

- upgrade'y - software, karty, łącza
- reorganizacja L2 – maksymalne ograniczenie rozciągania L2, zejście z VTP do poziomu lokalizacji
- readresacja L3: każda lokalizacja ma własne pule adresowe
- całkowite przejście z routingiem wewnętrznym na OSPF
- zejście z routingiem do poziomu klocków: agregacja rozmawia z modułami po OSPFie, to od modułów zależy który z nich podaje daną usługę (może podać ją także z innej lokalizacji, jeśli mamy do niej połączenie)

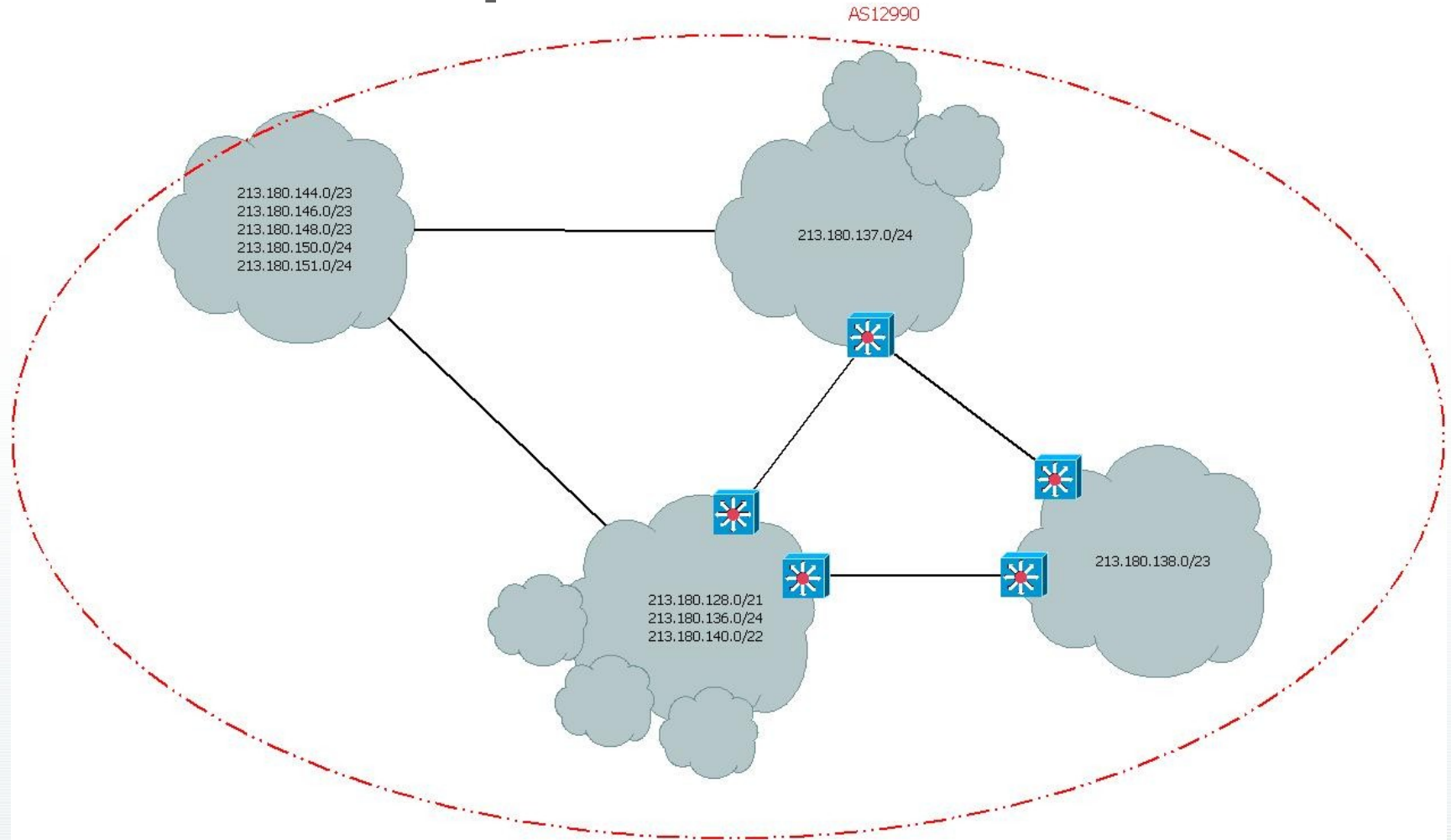


Zmiany 2008/2009 cd.

- rekonfiguracja BGP: konfederacje i route-reflectory
- rozgłaszanie do peerów BGP podsieci zamiast całej puli AS (w przypadku rozłączenia DCs każda lokalizacja będzie widziana w sieci osobno)
- pełne tablice BGP jedynie na routerach brzegowych
- testy niezawodności wdrażanych rozwiązań – symulacje awarii



BGP w Onet.pl SA



- pełen sukces ;)

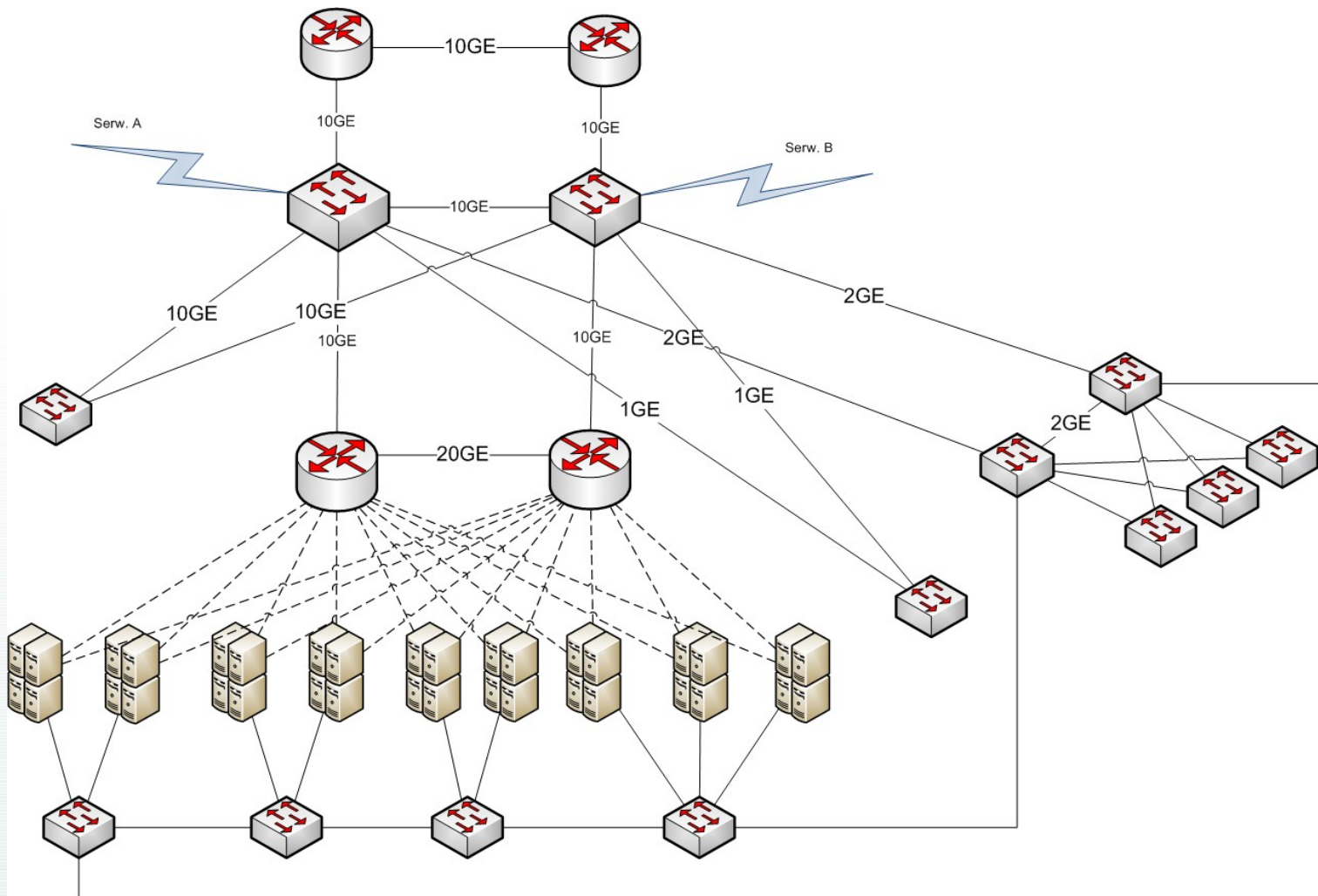


2009 Q2: startup ODC

- drugie co do wielkości Data Center w Polsce
- referencja Cisco jako pierwsza w Polsce kompletna architektura Data Center zbudowana na platformie Nexus
- w ODC współpracują ze sobą urządzenia sieciowe Cisco, Juniper, HP, Nortel, IBM...
- pod podłogami leży ponad 250 km kabli



ODC = 10GE



Agenda:

- Rys historyczny
- Rozwój i problemy sieci
- Nowa struktura
- Podsumowanie



W czym wiele DCs jest lepszych od 1 lub 2?

- 1 DC: najtaniej, ale spore ryzyko:
 - problem z modernizacjami w obiekcie
 - większy błąd ludzki – solidny downtime
 - utrata serwerowni to katastrofa
- 2 DCs: minimum bezpieczeństwa
 - łatwiej modernizować
 - pożar to spory problem, ale nie zostawia nas bez niczego
 - dobra konfiguracja pozwala skracać ewentualny downtime
 - możemy rozkładać zasoby i sterować zapełnieniem obiektów
- wiele DCs – podobne zalety jak 2 DCs, a dodatkowo:
 - lepsze rozłożenie zasobów (zapełnienie obiektów, dywersyfikacja usług)
 - odporność na trwałą niedostępność jednego DC
 - większa elastyczność w dostępie do operatorów



Checklista

- połączenia między DCs (kosztują)
- wykorzystanie potencjału kilku DCs na co dzień – architektura sieci oraz aplikacje, które są tego świadome
- plan na wypadek całkowitej utraty lokalizacji i danych w niej zgromadzonych
- obsługa serwerowni: połączenia, system przeciwpożarowy, dojazdy

ODC: zespół A1LW (16 osób), zdalnie A2LW (24 osoby)

Pozostałe serwerownie: bezobsługowe



Gdzie jesteśmy teraz?



- nawet na znakach drogowych ;)
- tankuje u nas content 12.5 miliona użytkowników w 3.5 mld odśłon miesięcznie
- serwujemy streaming, zarówno off-line jak i relacje na żywo
- podajemy na świat gigabity ruchu na sekundę - z trzech serwerowni, do kilkunastu peerów BGP
- przesyłamy dużo więcej danych wewnątrz korzystając z wydajnej, redundantnej struktury
- śpimy spokojniej ;)
- patrzymy w przyszłość



Pytania?



Dziękuję za uwagę.
Tadeusz.Knapik@portal.onet.pl

